

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:46:57 ; Search time 210.42 Seconds  
(without alignments)  
6.988 Million cell updates/sec

Title: US-09-331-631a-1\_COPY\_74\_116  
Perfect score: 248  
Sequence: 1 NQEDPQTECQCQCRRCRQDE.....RQOYQCRRCICEEEERY 43

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	248	100.0	666	19 W62828	Macadamia integrif
2	242	97.6	666	19 W62829	Macadamia integrif
3	241	97.2	625	19 W62830	Macadamia integrif
4	117	47.2	525	19 W62831	Theobroma cacao an
5	117	47.2	566	13 R20181	Sequence encoded b
6	109	44.0	590	19 W62832	Gossypium hirsutum
7	89	35.9	28	19 W62841	Stenocarpus sinuat
8	66.5	26.8	593	19 W62835	Zea mays antilicr
9	66.5	26.8	637	19 W62837	Hordeum vulgare an
10	65	26.2	919	10 P93109	Human androgen rec
11	65	26.2	919	18 W14783	Androgen receptor
12	65	26.2	919	21 Y78914	Human androgen rec

13	64.5	26.0	35	13 R21079	Antimicrobial maiz
14	63.5	25.6	33	19 W62836	Zea mays antimicro
15	63.5	25.6	51	18 W33694	Mouse protomine 1
16	63.5	25.6	176	18 W33695	Mouse protomine 1
17	63.5	25.6	301	19 W37085	Anti-human SC sing
18	63	25.4	910	20 Y22191	Mouse brain CNG-1
19	62.5	25.2	107	17 R91705	AcanAP23. Ancylos
20	62.5	25.2	107	20 Y30404	Nematode extracted
21	62.5	25.2	342	20 Y16785	Human secreted pro
22	60.5	24.4	215	14 R44806	Human cyclin D3 ps
23	60	24.2	154	20 Y33504	Human unliganded a
24	60	24.2	918	12 R12223	Human androgen rec
25	60	24.2	918	20 Y33491	Human androgen rec
26	59.5	24.0	98	21 Y65429	Human 5' EST relat
27	59	23.8	445	21 Y32374	Mouse CNRER-1. Mu
28	59	23.8	447	17 W03326	LXR-alpha, orphan
29	59	23.8	1447	20 W81029	Murine PCIP protei
30	58.5	23.6	88	20 Y30416	Mature nematode ex
31	58.5	23.6	514	19 W80400	A secreted protein
32	58	23.4	151	21 Y74634	Neisseria meningit
33	58	23.4	199	21 Y74635	Neisseria meningit
34	57.5	23.2	304	13 R20063	Human EGF-binding
35	57.5	23.2	317	10 P93396	Human follistatin
36	57.5	23.2	809	20 Y29672	Human cerebral pro
37	57	23.0	303	15 R60054	Dirofilaria immiti
38	57	23.0	440	14 R33744	XR2. Homo sapiens
39	57	23.0	567	20 Y22212	MTG16b protein seq
40	57	23.0	653	20 Y22211	MTG16a protein seq
41	57	23.0	704	20 Y22209	AML1-MTG8 fusion
42	57	23.0	752	15 R51701	AML1-MTG8 fusion
43	57	23.0	780	20 Y22210	AML1-MTG16 fusion
44	57	23.0	2023	21 Y54320	Amino acid sequenc
45	56.5	22.8	281	21 Y91958	Human cytoskeleton

## ALIGNMENTS

RESULT 1	
W62828	W62828 standard; Protein; 666 AA.
XX	
AC	W62828;
XX	
DT	27-OCR-1998 (first entry)
XX	
DE	Macadamia integrifolia antimicrobial protein.
XX	
KW	antimicrobial protein; infestation; control.
XX	
OS	Macadamia integrifolia.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..28
FT	/note= "signal peptide"
FT	29..666
FT	Protein
XX	/note= "mature protein"
XX	
PN	W09827805-A1.
XX	
PD	02-JUL-1998.
XX	
PF	22-DEC-1997; 97WO-AU00874.
XX	
PR	20-DEC-1996; 96AU-0004275.
XX	
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX	
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX	WPI; 1998-377279/32.
DR	N-PSDB; V42310.
XX	

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 34-36; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 100.0%; Score 248; DB 19; Length 666;  
Best Local Similarity 100.0%; Pred. No. 5.8e-20;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NOEDPOTECQCCORRCROESGPRQOQYCORCKEICEEEY 43  
Db 74 ngddpqlcqcqqrccrgqsgsprqgycqrckelceeeey 116

RESULT 2  
ID W62829 standard; Protein; 666 AA.  
XX W62829;  
XX  
DT 27-OCT-1998 (first entry)  
XX  
DE Macadamia integrifolia antimicrobial protein.  
XX  
KM antimicrobial protein; infestation; control.  
XX  
OS Macadamia integrifolia.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..28 /note= "signal peptide"  
FT Protein 29..666 /note= "mature protein"  
XX  
PN MO9827805-A1.  
XX  
PD 02-JUL-1998.  
XX  
PE 22-DEC-1997; 97WO-AU00874.  
XX  
PR 20-DEC-1996; 96AU-0004275.  
XX  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
XX  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
XX  
DR WPI; 1998-377279/32.  
XX  
DR N-PSDB; V42311.  
XX  
XX Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 39-41; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 666 AA;

Query Match 97.6%; Score 242; DB 19; Length 666;  
Best Local Similarity 95.3%; Pred. No. 2.7e-19;  
Matches 41; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NOEDPOTECQCCORRCROESGPRQOQYCORCKEICEEEY 43

Db 74 ngddpqlcqcqqrccrgqsgsprqgycqrckelceeeey 116  
||:||||:|||||

RESULT 3  
ID W62830 standard; Protein; 625 AA.  
XX W62830;  
XX  
DT 27-OCT-1998 (first entry)  
XX  
DE Macadamia integrifolia antimicrobial protein.  
XX  
KM antimicrobial protein; infestation; control.  
XX  
OS Macadamia integrifolia.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..28 /note= "signal peptide"  
FT Protein 29..666 /note= "mature protein"  
XX  
PN MO9827805-A1.  
XX  
PD 02-JUL-1998.  
XX  
PE 22-DEC-1997; 97WO-AU00874.  
XX  
PR 20-DEC-1996; 96AU-0004275.  
XX  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
XX  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
XX  
DR WPI; 1998-377279/32.  
XX  
DR N-PSDB; V42316.  
XX  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 43-45; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 625 AA;

Query Match 97.2%; Score 241; DB 19; Length 625;  
Best Local Similarity 97.7%; Pred. No. 3.2e-19;  
Matches 42; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NOEDPOTECQCCORRCROESGPRQOQYCORCKEICEEEY 43  
Db 33 ngddpqlcqcqqrccrgqsgsprqgycqrckelceeeey 75

RESULT 4  
ID W62831 standard; Protein; 525 AA.  
XX W62831;  
XX  
DT 27-OCT-1998 (first entry)  
XX  
DE Theobroma cacao antimicrobial protein.  
XX  
KM antimicrobial protein; infestation; control.  
XX  
OS Theobroma cacao.

xx	MO9827805-A1.
pn	
xx	
xx	02-JUL-1998.
pd	
xx	
xx	22-DEC-1997; 97WO-AU00874.
pf	
xx	
xx	20-DEC-1996; 96AU-0004275.
pr	
xx	
pa	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
xx	
pi	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
xx	
dr	WPI; 1998-377279/32.
xx	
xx	Novel anti-microbial protein from e.g. <i>Macadamia integrifolia</i> -
pt	useful for controlling microbial infestations of plants or mammals
xx	
ps	Claim 1; Page 47-49; 96pp; English.
xx	
xx	
cc	The sequence is that of an antimicrobial protein which can
cc	be used to control microbial infestations in plants and mammalian
cc	animals.
xx	
xx	
50	Sequence 525 AA;

	Query Match	47.2%	Score 117	DB 19	Length 525
	Best Local Similarity	50.0%	Pred. No. 1.2e-05		
	Matches 20	Conservative 10	Mismatches 10	Indels	Gaps 0
QY	2 QEDPTECCGCCRCRQDSGPRDQOYCCRCRKEICEEBE	41			
db	78 eedlqrqyqccqgricqeqqgqqrqeqqqrkweyqkege	117			

RESULT	5
R20181	
ID	R20181 standard; Protein; 566 AA.
XX	
AC	R20181;
XX	
DT	16-APR-1992 (first entry)
XX	
DE	Sequence encoded by 67 kD T. cacao protein cDNA.
XX	
KW	Cocoa; flavour; vicilin; seed storage protein.
XX	
OS	Theobroma cacao.
XX	
PN	WO9119801-A.
XX	
PD	26-DEC-1991.
XX	
PF	07-JUN-1991; 91WO-GB00914.
XX	
PR	11-JUN-1990; 90GB-0013016.
XX	
PA	(MRSC ) MARS UK LTD.
XX	
PI	Spencer ME, Hodge R, Deakin EA, Ashton S;
XX	
DR	WPI; 1992-024418/03.
RR	N-PDB; Q20377.
XX	
PT	Recombinant cocoa proteins - are responsible for flavour in coccoa
PT	beans and produced in large quantities using yeast and bacterial
PT	expression vectors
XX	
PS	Claim 4; Fig 2; 59pp; English.
XX	
CC	The inventors claim a 67 kD and 31 kD T. cacao protein, and
CC	fragments, and encoding DNAs. The 47 kD and 31 kD proteins are

CC derived from the 67 kD precursor. <sup>1</sup> cacao protein cDNA was  
CC detected in a cDNA library prepared from immature cocoa beans RNA  
CC using a probe based on the AA sequence of a CNR peptide common to  
CC the 47 kD and 31 kD polypeptides. Homology searches revealed close  
CC homologies between the 67 kD polypeptide and the vicilins, which are  
CC seed storage proteins.

Query Match	47.2%	Score 117	DB 13	Length 566
Best Local Similarity	50.0%	Pred No. 1.3e-05		
Matches	20	Conservative	10	Mismatches 10
				Indels 0
				Gaps 0

```

QY 2 QELPQTECCQCCQRCRCRQDESGPPQOQYQRCRCRCEICEEEE 41
    :|: | : ||| ||: : | :| ||: | : |:|
Db 78 eeelqrgyqqcgrlrcgeqqgqgqeqgcqrkckeqykege 117

```

RESULT	6
W62832	
ID	W62832 standard; Protein; 590 AA.
XX	W62832;
XX	
DT	27-OCT-1998 (first entry)

DE *Gossypium hirsutum* antimicrobial protein.  
XX  
KW antimicrobial protein; infestation; control

OS *Gossypium hirsutum*.

PN	W09827805-A1
XX	
PD	02-JUL-1998.

PF	22-DEC-1997;	97WO-AU00874.
XX		
PR	20-DEC-1996;	96AU-0004275.

PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
XX  
XX  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP,  
XX  
XX  
DR WPI: 1998-377279/32.

PT Novel anti-microbial protein from e.g. *Macadamia integrifolia* -  
PT useful for controlling microbial infestations of plants or mammals  
XX  
PS Claim 1; Page 49-51; 96pp; English.

CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.

SQ Sequence 590 AA;

Query Match	44.0%	Score 109;	DB 19;	Length 590;
Best Local Similarity	50.0%	Pred. No. 0.0001;		
Matches 20;	Conservative	8;	Mismatches 10;	Indels 2;
				Gaps 1;

```

QY.      3 EDPQTECCCCRRRCROESGPRQOQYCORCKETCEEEEE 42
          |||  ::||  |||  |||  |||  :  |||  :
Db      81 eDPqrrYeeCqgeCrgeee--rqpqCqqrC|krfegqg 118

```

RESULT	7
W62841	
ID	W62841 standard; Protein; 28 AA.
XX	
AC	W62841;

```

XX 27-OCT-1998 (first entry)
XX
XX Stenocarpus sinuatus antimicrobial protein.
XX DE antimicrobial protein; infestation; control.
XX KW
XX Stenocarpus sinuatus.
XX OS
XX MO9827805-A1.
XX PN
XX 02-JUL-1998.
XX PD
XX 22-DEC-1997; 97MO-AU00874.
XX PE
XX 20-DEC-1996; 96AU-0004275.
XX PR
XX (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX PA
XX Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP,
XX PI
XX MPI: 1998-377279/32.
XX DR
XX
XX Novel anti-microbial protein from e.g. Macadamia integrifolia -
XX PT useful for controlling microbial infestations of plants or mammals
XX
XX Claim 1; Page 66; 96pp; English.
XX PS
XX
XX The sequence is that of an antimicrobial protein which can
XX CC be used to control microbial infestations in plants and mammalian
XX CC animals.
XX
XX Sequence 28 AA:
XX
Query Match 35.9%; Score 89; DB 19; Length 28;
Best Local Similarity 59.3%; Pred. No. 0.00093;
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0.
07 4 DPQTECQCQQRRCRQESGPRQDTCQ 30
07 11 : 1 1 1 1 1 1 1 1 1 1 1 1 :
db 2 dplrqqlcmrcqgqekdprqgqcx 28

```

RESULT	8
XX	W62835
XX	W62835 standard; Protein; 593 AA.
XX	AC
XX	W62835;
XX	27-OCT-1998 (first entry)
XX	DE
XX	Zea mays antimicrobial protein.
XX	antimicrobial protein; infestation; control.
XX	OS
XX	Zea mays.
XX	W09827805-A1.
XX	PD
XX	02-JUL-1998.
XX	PP
XX	22-DEC-1997; 97MO-AU00874.
XX	PR
XX	20-DEC-1996; 96AU-0004275.
XX	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX	PI
XX	Bower NI, Goulter KC, Green JT, Manners JM, Marcus JP;
XX	WPI; 1998-377279/32.
XX	Novel anti-microbial protein from e.g. Macadamia integrifolia -

[illegible]

```

OY      2 QEDPQTECCQCORRC-RQESGPRQQYCCQRCK 34  
       :|::||:|::|::|::|:  
Db      557 eeersgrgcrrgcllrheggpwetgecmrccr 590  
  
RESULT          9  
ID W62837 standard; Protein; 637 AA.  
AC W62837;  
XX  
DF 27-OCT-1998 (first entry)  
DE Hordeum vulgare antimicrobial protein.  
XX  
KM antimicrobial protein; Infestation; control.  
XX  
OS Hordeum vulgare.  
PN W09827805-A1.  
XX  
PD 02-JUL-1998.  
XX  
PE 22-DEC-1997; 97MO-AU00874.  
XX  
PR 20-DEC-1996; 96AU-0004275.  
XX  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
XX  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
XX  
DR WPI: 1998-377279/32.  
XX  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
XX useful for controlling microbial infestations of plants or mammals  
PS Claim 1; Page 60-62; 96pp; English.  
XX  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
XX  
SQ Sequence 637 AA;  
  
Query Match 26.8%; Score 66.5; DB 19; Length 637;  
Best Local Similarity 39.4%; Pred. No. 5.3;  
Matches 13; Conservative 7; Mismatches 8; Indels 5; Gaps 2;  
OY 10 QQCORCRHQESGPRQQYCCQRCKEICEEEE 42  
   |||:||||: | | | | |:  
Db 42 qgcvcrcrgcr--pr---ysharcvgeordqg 69  
  
RESULT 10  
ID P93109 standard; protein; 919 AA.  
XX
```







